

Acceptor site, datasæt

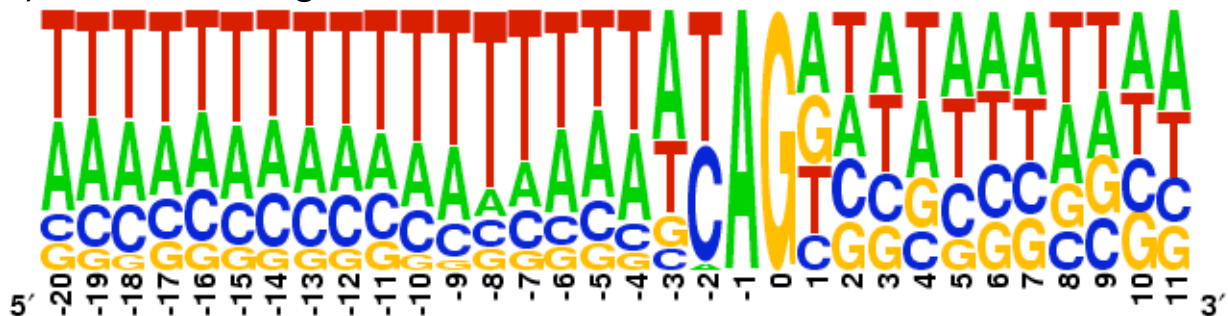
- Problemstilling: Hvad er det biologiske signal omkring acceptor-site?
- Datasæt: 268 acceptor sites fra Gær.
- Følgende side viser hvordan **informationen** i hyppigheden af nukleotid-forekomsten kan behandles.

... **AG**

Intron	Exon
GTTCTTCGTGTTTATTTT	AGGAAATTGATGA
TTGTTTCTCCTTTTAA	ATAGTACTGCTGTTT
TTTACTAACGACACATT	GAAATCACTTTG
GATACGCTTACCGTTAT	CCAGAGCTACAGCGC
TACTAATATGTAATACT	TCAGCTCCCCTTAAT
ATTGAGATCTTTTTTAA	CTAGTTAGGTCTACC
TTCTCCCCTTCTTCATT	TTTAGCCTGTTTGGAC
TAACATAACTTATTTAC	ATAGTGCCATTGAAC
GATATTTCCCGTTGTGT	TAAGGCTGAGAAGAA
TTTTCCCGACCATCAAG	ACAGGTGATTTATCA
TGCAAAAACTTTTTTTC	ACAGGGCTAACTTGC
GTTTATTGTGTTTCCAC	TCACTTAAAAACGA
AACGTACTTTAATATTT	TATAGTACTTCATTCC
AACATGCTATTTTTTC	ATACAGCAACCTCACAT
CTGCACTCATCATTAG	ATTAGAGGAACATGGA
TACTTTTCTTTATCTA	AGCAGCTAACTCAACT
ATCAACATGCTATTGAA	CTAGAGATCCACCTA
TAACATGACTTTAACAG	GGGCTAATTTTAC
AGTACTAACTAATTAAC	TAGAACATTAACAT
GATCACCCTCACATTAT	TAGAAATTTCAAACG
CAGTGGAATTTTTTTT	CTAGAAATGGTATCG
CTCTATGACCAATAAAA	ACAGACTGTACTTTTC
AAATGGTATTATTTATA	ACAGTTGAACATTTTC
ATAAATATGCGATCAAT	ATAGACCGTTGATAT
ATTTTACTTTTTTTTTT	TAGGAGCTCCAAGA
ATTTATTTCCCTTATAA	TACAGACACGGTTACA
TCGCAATTAATTTTCTA	ATAGTTTTTCATTTT
GACCATCTTTCTTTTCC	CCAGTGCTAAACACG
AACCTTCTTTCTCATTC	GTAGATTACTGTTGC
AATTACTAACAGCTGTA	ATAGCCGACAAATTT
CTCTCTGCGCGTCCAAT	TAGCTATACTGTTG
TTGTTTTGTTTTGTCTA	CGTGTGTTGGAGAA
AACTTCCATTTCTTACA	TAGATCATCGCCAT
TCCTTTCCATAATTTAT	TACAGCGCTTTGGTAT
CGATTTACTATTTCCATT	TAGACGTTGTTCAA
AATTTACTAACAATACT	TCAGTTTATAATGGA
TCCTATACTAACAATTT	GTAGTTCATAAATAA
...	...

A	94	88	84	75	78	78	71	69	70	60	68	77	32	49	87	93	93	134	9	266	0	86	66	85	81	89	81	88	82
C	31	45	52	44	56	46	62	54	56	51	46	37	30	42	32	44	30	25	122	1	0	38	65	52	43	62	62	57	43
T	113	110	113	117	104	117	111	120	118	125	136	140	182	155	122	100	124	75	137	0	0	72	85	82	91	83	73	67	96
G	30	25	19	32	30	27	24	25	24	32	18	14	24	22	27	31	21	34	0	1	268	72	52	49	53	34	52	56	47

A	0,35	0,33	0,31	0,28	0,29	0,29	0,26	0,26	0,26	0,22	0,25	0,29	0,12	0,18	0,32	0,35	0,35	0,50	0,03	0,99	0,00	0,32	0,25	0,32	0,30	0,33	0,30	0,33	0,31
C	0,12	0,17	0,19	0,16	0,21	0,17	0,23	0,20	0,21	0,19	0,17	0,14	0,11	0,16	0,12	0,16	0,11	0,09	0,46	0,00	0,00	0,14	0,24	0,19	0,16	0,23	0,23	0,21	0,16
T	0,42	0,41	0,42	0,44	0,39	0,44	0,41	0,45	0,44	0,47	0,51	0,52	0,68	0,58	0,46	0,37	0,46	0,28	0,51	0,00	0,00	0,27	0,32	0,31	0,34	0,31	0,27	0,25	0,36
G	0,11	0,09	0,07	0,12	0,11	0,10	0,09	0,09	0,09	0,12	0,07	0,05	0,09	0,08	0,10	0,12	0,08	0,13	0,00	0,00	1,00	0,27	0,19	0,18	0,20	0,13	0,19	0,21	0,18



Informations beregning:

$$R_{seq} = S_{max} - S_{obs} = \log_2 N - \left(- \sum_{n=1}^N p_n \log_2 p_n \right)$$

